



Segmentez des clients d'un site e-commerce

Aïcha HACHEMI

OpenClassroom

SOMMAIRE

Présentation du projet.

Analyse Exploratoire.

Segmentation RFM.

Segmentation RFM et d'autres variables.

Choix du nombre de clusters.

Stabilité temporelle et maintenance.





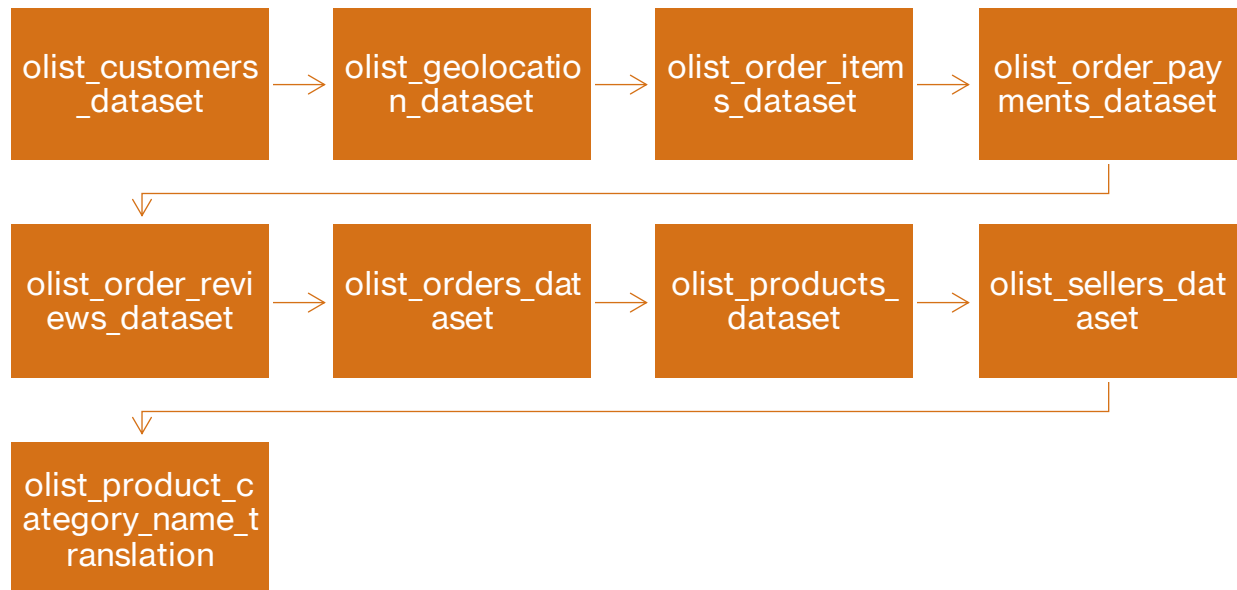
Présentation du projet

Mission

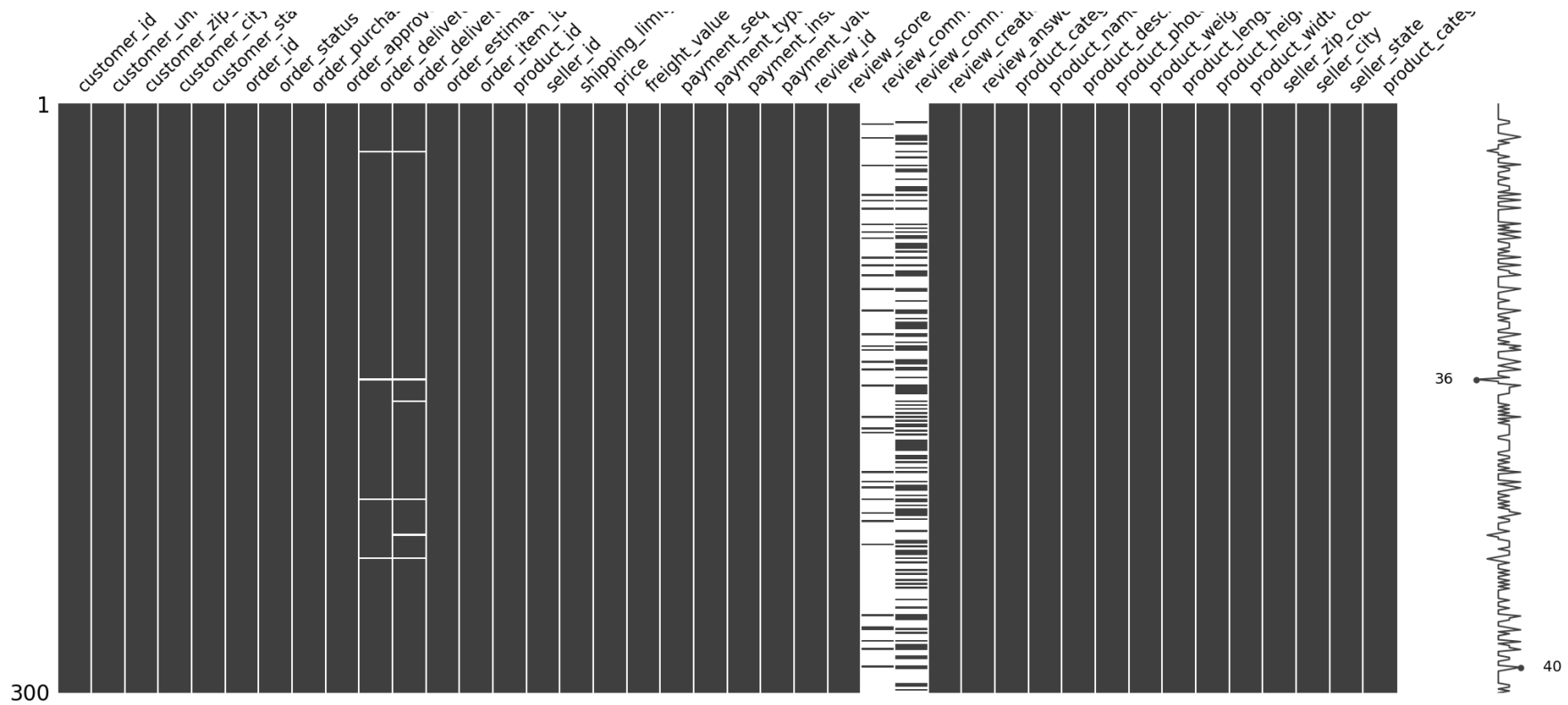
Olist, une entreprise brésilienne qui propose une solution de vente sur les marketplaces en ligne.

Fournir à ses équipes d'e-commerce une **segmentation des clients** qu'elles pourront utiliser au quotidien pour leurs campagnes de communication.

Analyse Exploratoire

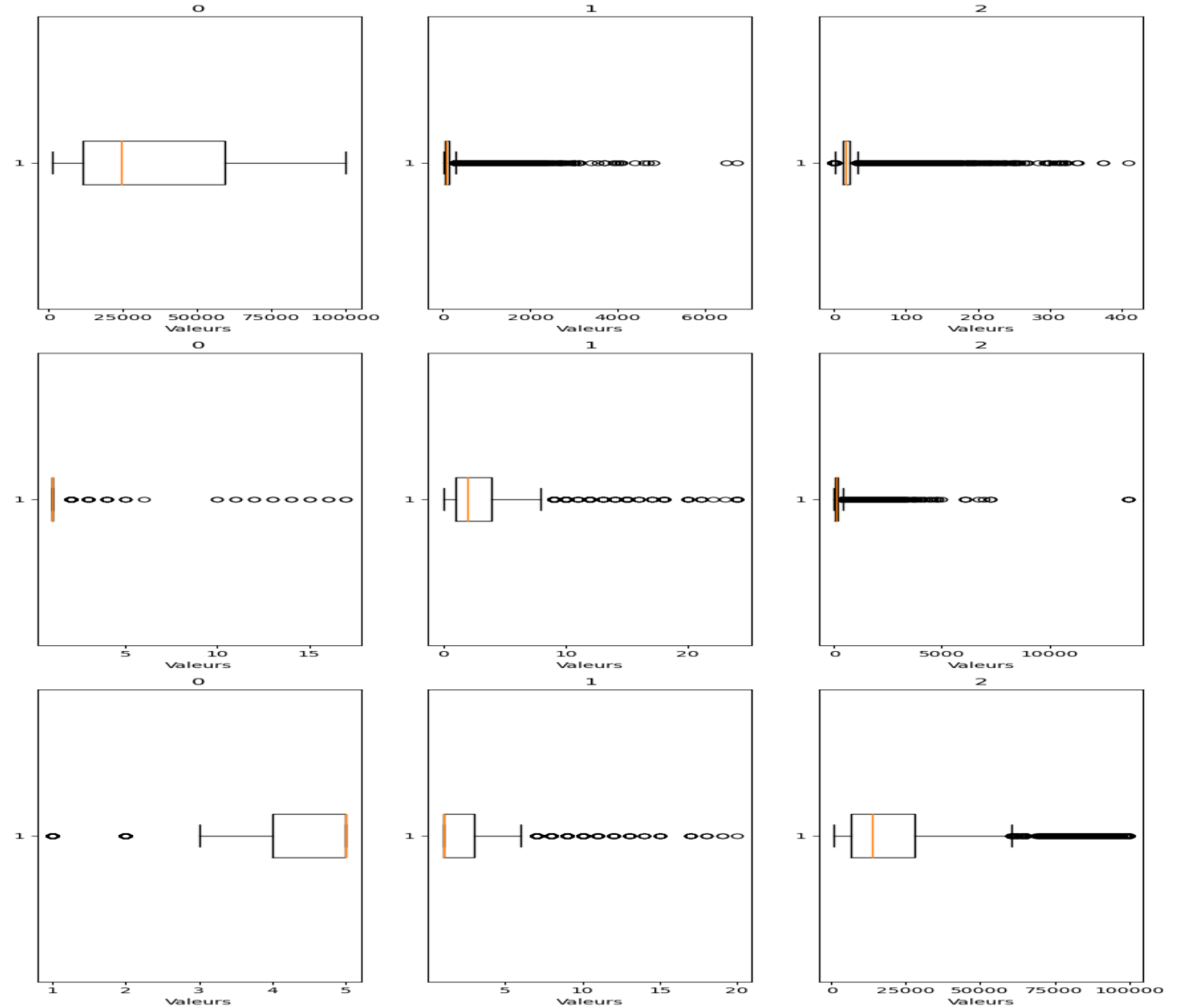


Analyse Exploratoire





Analyse Exploratoire



	Recency	Frequency	MonetaryValue
customer_unique_id			
4222f488432f87a3cf7a836582378acc	539	3	217.74
b8b6643025dbcaa0e516a72f6d0d0717	388	2	244.44
c2c7712d6c0a2ac5be6a3280042d471b	464	2	192.60

Segmentation RFM

Segmentation RFM

Segmentation basée sur les quartiles

Segmentation avec l'algorithme k-means

Segmentation RFM

Quartiles de scorings

Recency Frequency MonetaryValue r_score f_score m_score rfm_score

customer_unique_id

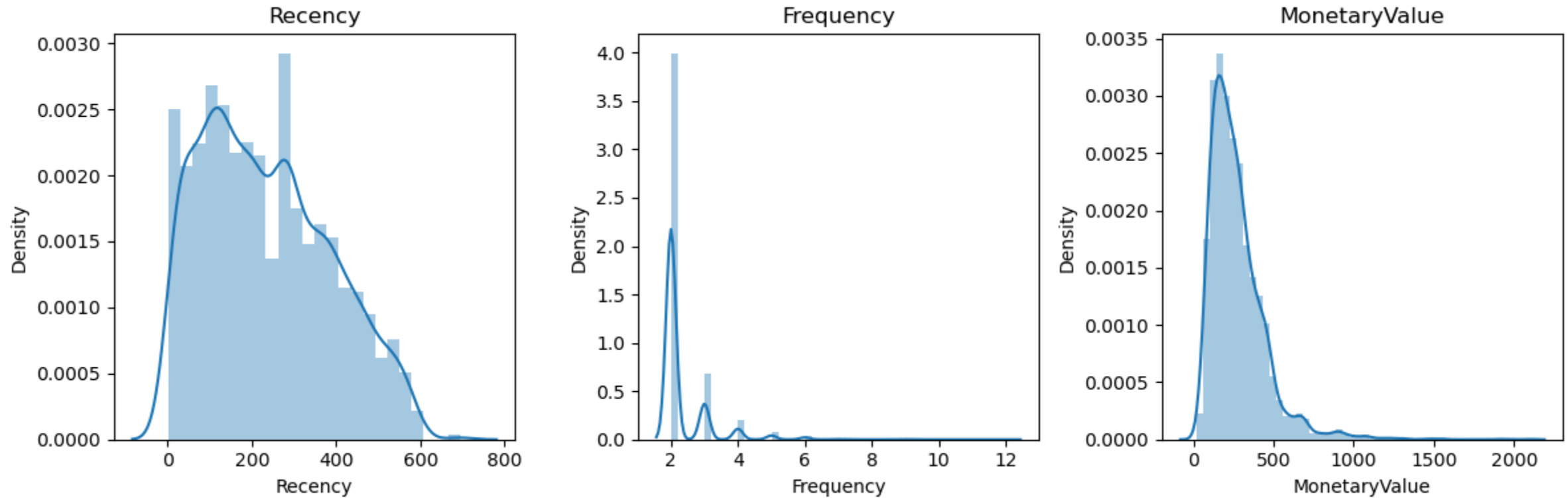
000bfa1d2f1a41876493be685390d6d3	335	2	93.70	2	1	1	211
001147e649a7b1afd577e873841632dd	364	2	424.32	1	1	4	114
003e2af81ea3c72fbdd093f07724a4de	136	2	127.48	3	1	1	311
004f84d065096f3cedd26cf0adc80b9d	126	2	378.16	3	1	4	314
00504b68faa074d0f40adcb4f7727f71	18	2	397.40	4	1	4	414

Recency Frequency MonetaryValue
mean mean mean count

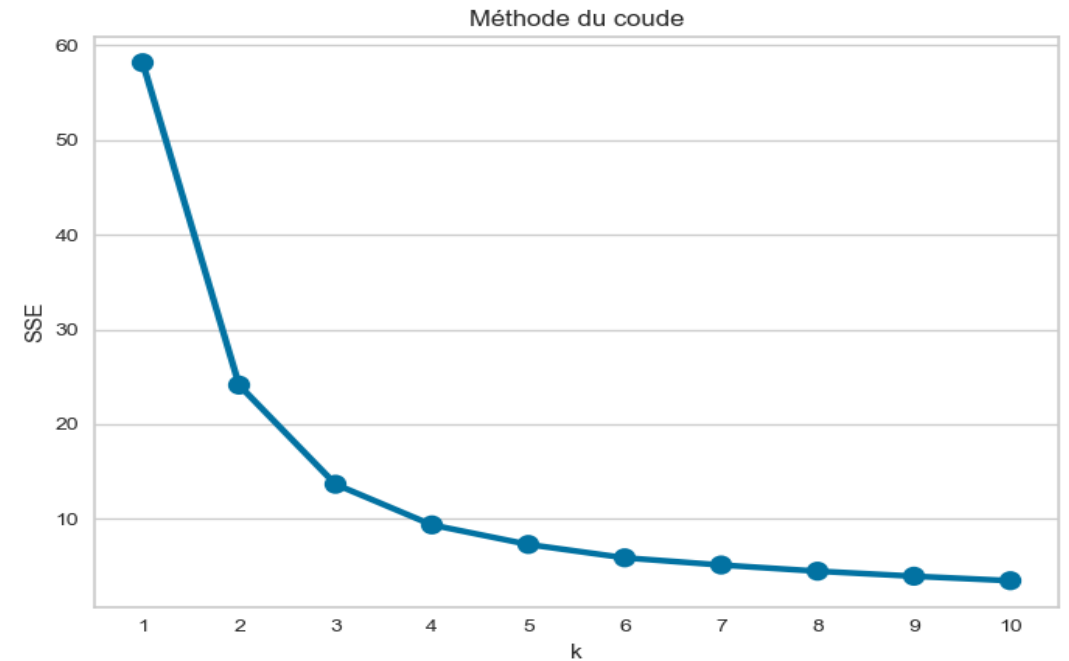
General_Segment

Gold	56.2	2.3	271.1	1003
Silver	160.1	2.3	272.9	1000
Bronze	360.8	2.3	275.3	2002

Segmentation RFM-preprocessing



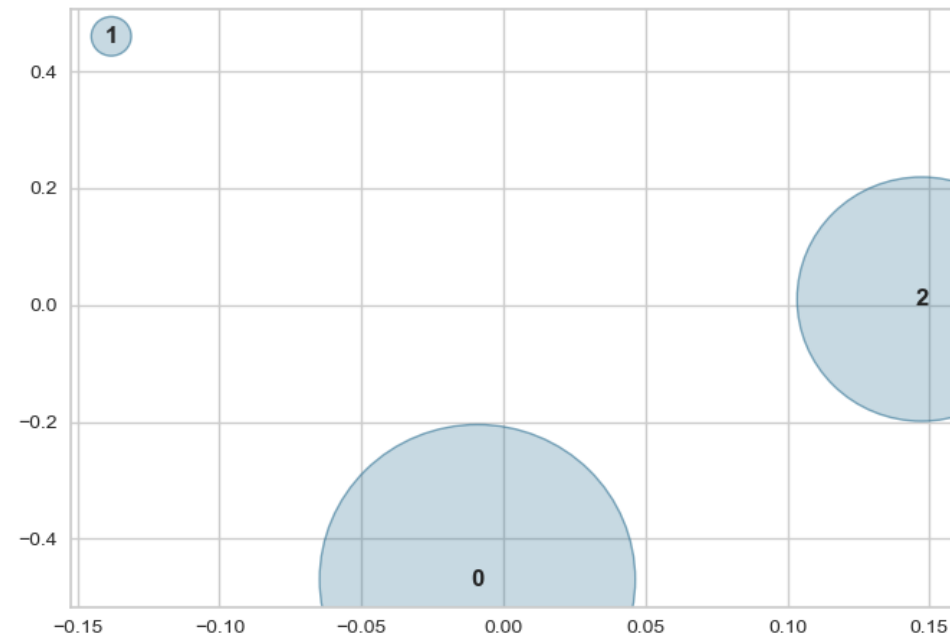
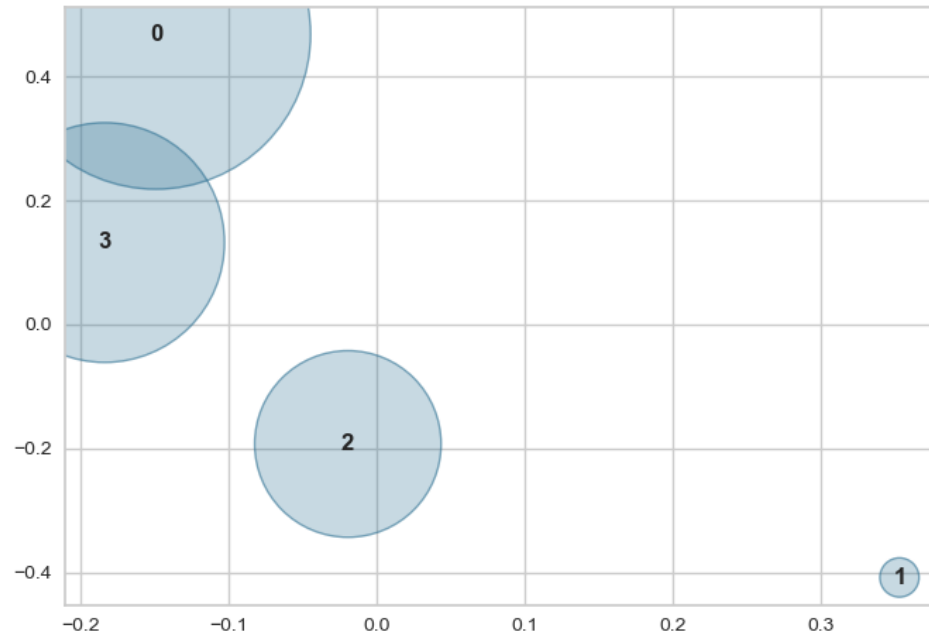
Segmenation RFM Algor ithme k-means



Pour `n_clusters=2`, le silhouette score est 0.5670580936489371
Pour `n_clusters=3`, le silhouette score est 0.48315972577297617
Pour `n_clusters=4`, le silhouette score est 0.45067924735826437
Pour `n_clusters=5`, le silhouette score est 0.4306696492031905
Pour `n_clusters=6`, le silhouette score est 0.4216776903778302
Pour `n_clusters=7`, le silhouette score est 0.43414421529941843
Pour `n_clusters=8`, le silhouette score est 0.4131502552352538

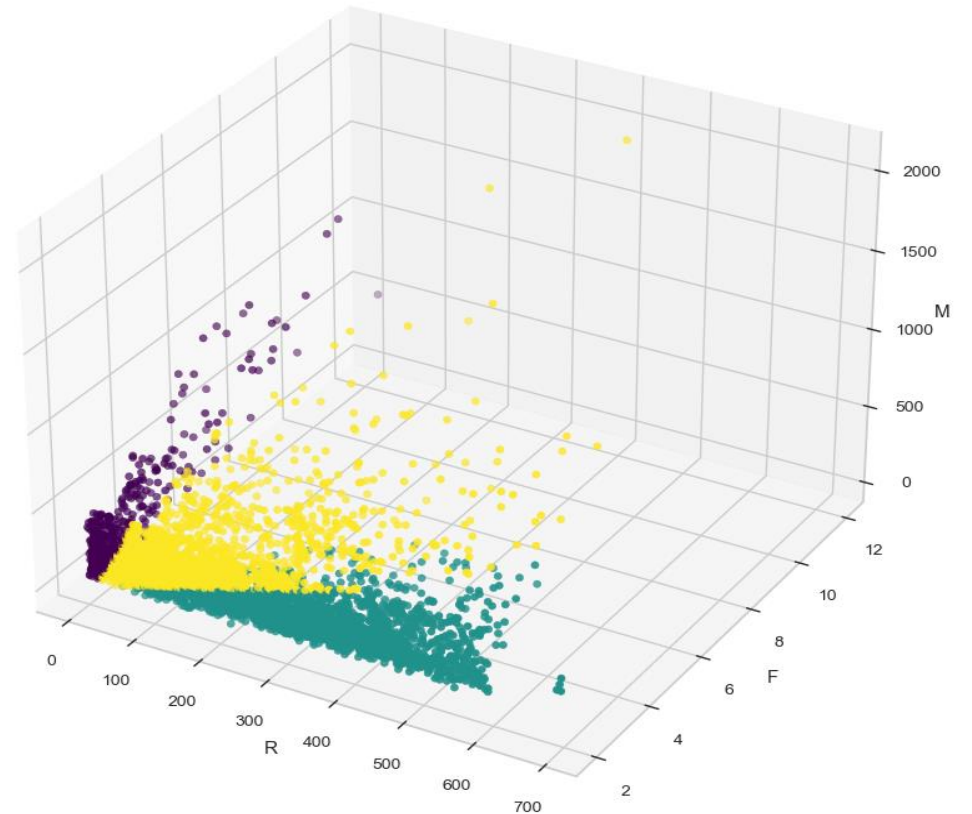
Segmenation RFM

Algorithme k-means



Segmenation RFM

Algorithme k-
means



Segmentation RFM et d'autres variables

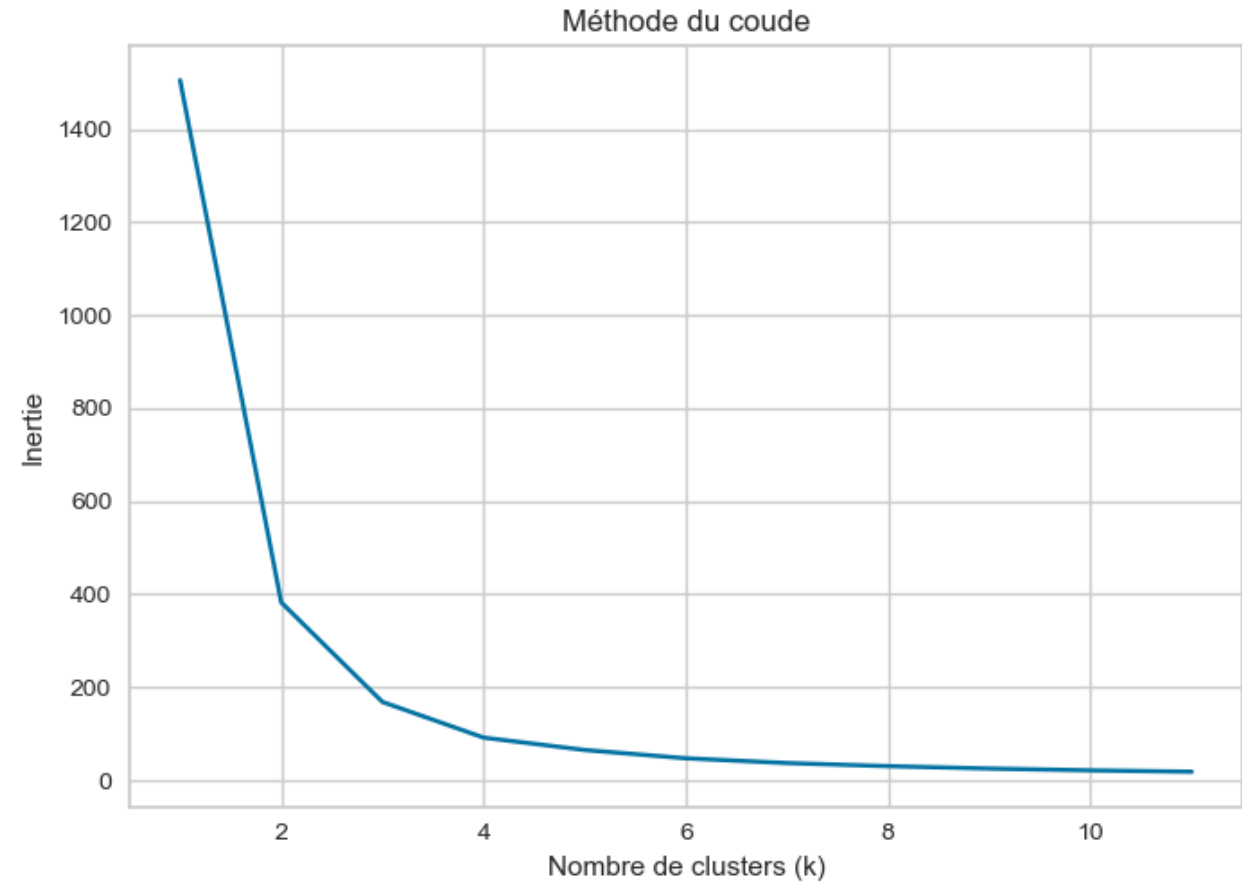
customer_zip_code_prefix'

'price'

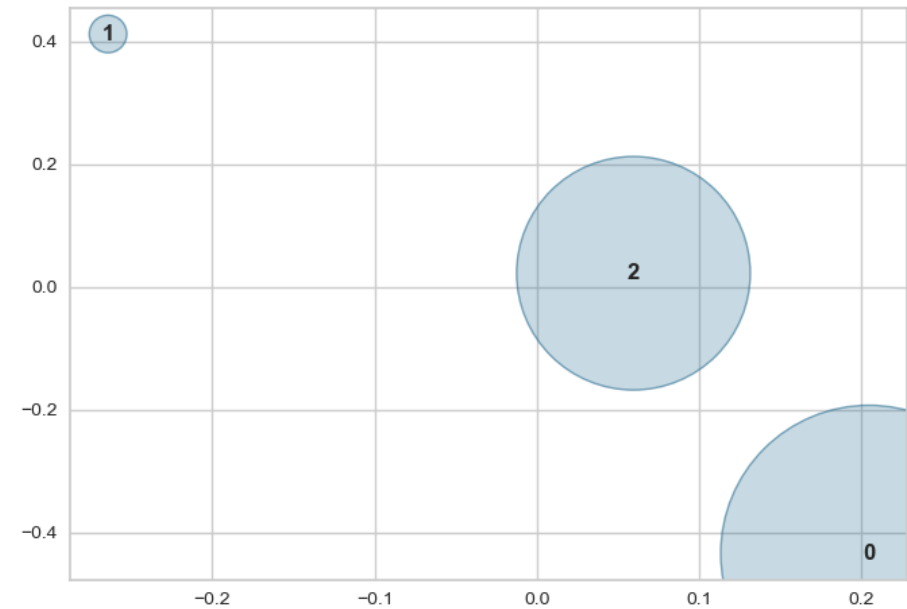
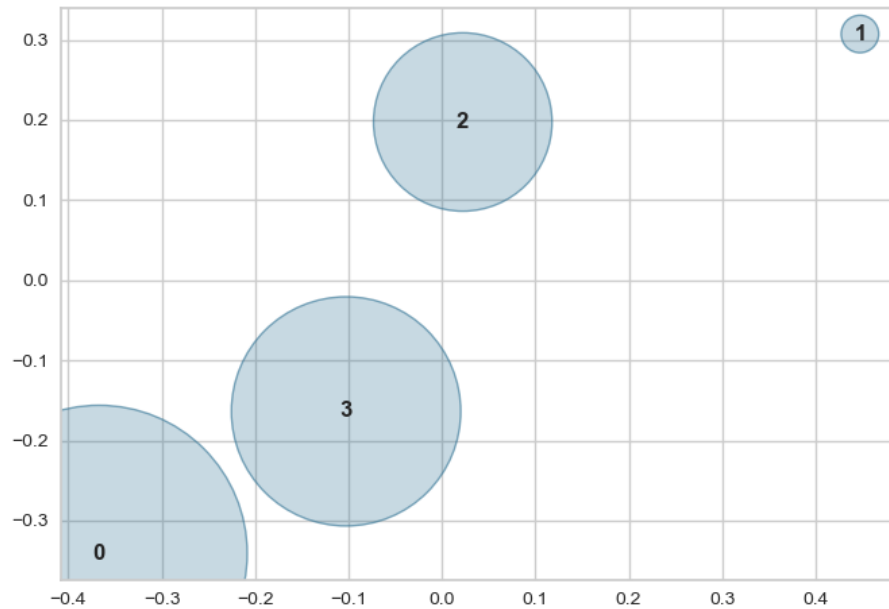
'review_score'

'seller_zip_code_prefix'

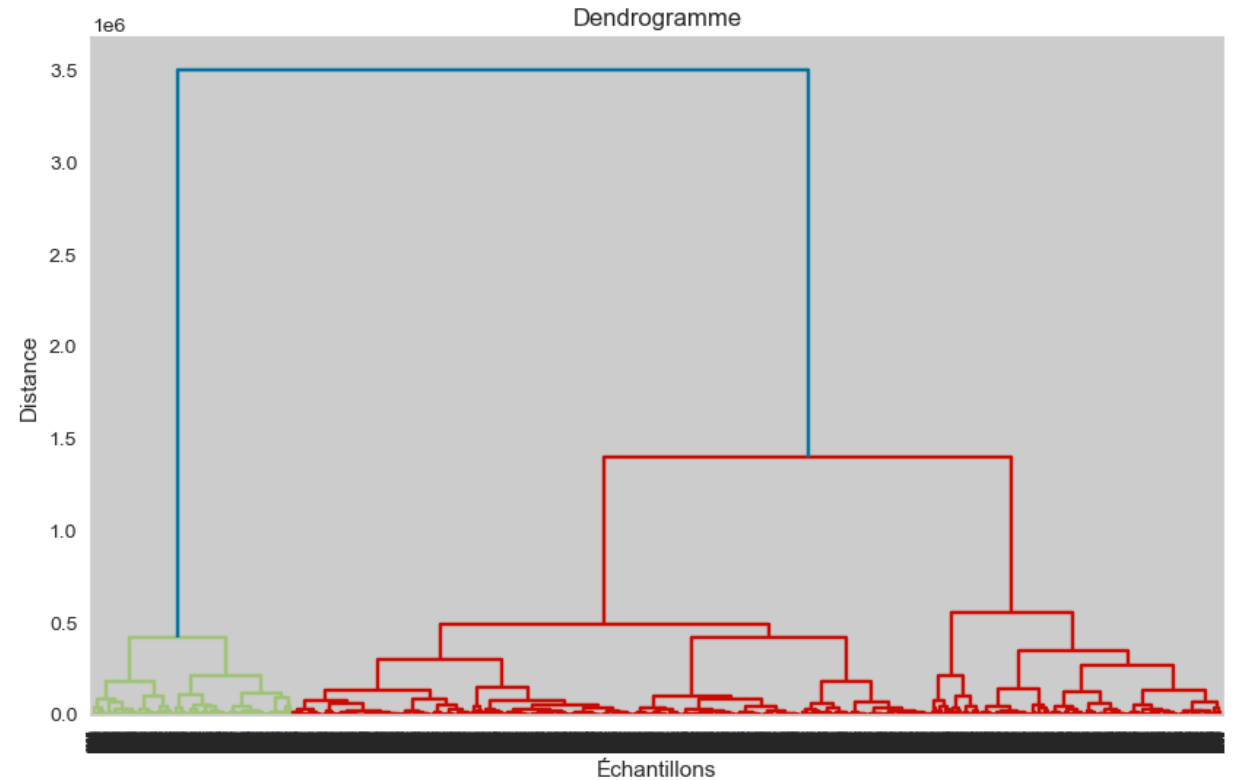
Segmentation RFM et d'autres variables



Segmentation RFM et d'autres variables



Segmentation RFM et d'autres variables



Maintenance

ARI entre les clusters à T0 et T1 :
0.8112408354280909

ARI entre les clusters à T0 et T2 :
0.4556358156127056

Conclusion : maintenance tous les
six mois.

